Digital Object Identifier 10.1109/ACCESS.2017.DOI

# Cooperative Beamforming with Predictive Relay Selection for Urban mmWave Communications

## ANASTASIOS DIMAS<sup>1</sup>, (Student Member, IEEE), DIONYSIOS S. KALOGERIAS<sup>2</sup>, (Member, IEEE), and ATHINA PETROPULU<sup>1</sup>, (Fellow, IEEE)

<sup>1</sup>Department of Electrical and Computer Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA (e-mail: tasos.dimas@rutgers.edu; athinap@rutgers.edu)

<sup>2</sup>Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA (e-mail: dionysis@seas.upenn.edu) Corresponding author: Anastasios Dimas (e-mail: tasos.dimas@rutgers.edu).

This work was supported by the National Science Foundation under Grant CCF-1526908.

### ABSTRACT

While millimeter wave (mmWave) communications promise high data rates, their sensitivity to blockage and severe signal attenuation present challenges in their deployment in urban settings. One effective way to mitigate those effects and increase communication range is relaying in combination with beamforming. However, relay-assisted beamforming also requires relay selection, a process that not only is resource demanding due to the need for explicit Channel State Information (CSI), but also introduces network latency. In this paper, we introduce a relay beamforming approach for mmWave communications in an urban scenario, which exploits the shadowing-induced correlation structure of the channel to reduce both latency and CSI estimation overhead. Our system model consists of static relays deployed in clusters across streets. Under that setting, we propose a novel, resource efficient scheme for joint optimal relay selection and distributed cooperative beamforming that maximizes the expected Signal-to-Interference plus Noise Ratio (SINR) at the destination, under power constraints. In particular, assuming a time-slotted system operation, one relay from each cluster is optimally selected at each time slot to participate in optimal beamforming at the next time slot. The key novelty of the proposed scheme is that relay selection is implemented in a predictive and distributed manner, by exploiting channel correlations and by using past and present measurements of magnitude CSI. As a result, at each time slot, optimal beamforming based on relays selected in the previous slot and optimal predictive relay selection for the next time slot are implemented completely in parallel. This parallelization eliminates delays induced by sequential execution of relay selection and beamforming, and substantially reduces CSI estimation overhead. Simulations confirm that the proposed relay selection scheme outperforms any randomized selection policy, while, at the same time, achieves comparable performance to an ideal selection scheme that relies on perfect CSI estimates for all candidate relays.

**INDEX TERMS** mmWave, cooperative relay beamforming, relay selection, 2-hop relaying, mmWave channel modeling, Vehicle-to-Infrastructure.

#### I. INTRODUCTION

T HE continuously growing number of connected devices has led to congestion in the licensed spectrum. To alleviate the problem, next generation of commercial wireless networks will exploit the previously untapped millimeter wave (mmWave) spectrum band [1]. The mmWave spectrum encompasses the frequencies 30 - 300 GHz, however the considered mmWave bands for use in urban cellular systems are at 28 and 38 GHz, since it has been shown that rain attenuation and oxygen absorption is low [2], [3] at those frequencies. The large available bandwidth will enable data rates of the order of Gigabits-per-second (Gbps), and thus will be able to support future wireless applications such as Vehicle-to-Vehicle (V2V) or Vehicle-to-Infrastructure (V2I) communications. However, to exploit the full potential of mmWave networks for outdoor-to-outdoor wireless communications, a new set of challenges, which are not present in current wireless networks would need to be overcome, such as severe path-loss, and sensitivity to blockage, shadowing and fading [4].

Various measurement campaigns [2], [5]–[7] have revealed that mmWaves incur increased propagation path-loss as compared to the sub-6 GHz frequencies used today. One way to compensate for that effect is to employ transmit beamforming [8], which can increase the probability of the signal arriving successfully at the destination. MmWaves also experience large-scale fading, which, according to measurement campaigns conducted in New York City and Austin, Texas [2], can be modeled as normally distributed in the dB domain. A similar campaign conducted in Daejeon City, Korea led to similar conclusions [5]. MmWave wireless communications in densely built cities are also susceptible to blockage by their surrounding environment, e.g., by high-rise buildings, moving vehicles, pedestrians, etc. [9]-[11]. In such an environment, the dominant propagation pattern of mmWaves is along street canyons by reflecting off of buildings; this is facilitated by the high reflectivity of common materials found in cities, such as concrete and glass [12].

One way to overcome blockage effects would be to deploy a dense network of street level mmWave base stations. Although this would optimize the city wide Line-of-Sight (LoS) coverage rate [13], it would not provide Quality-of-Service (QoS) assurances. In fact, a dense network of mmWave base stations can degrade the QoS of users due to interfering base station transmissions [14], [15]. An alternative solution to overcoming blockage and improve coverage is through relaying [4], [8], [16], a concept that has been included in the currently applied LTE-Advanced standard [17]. Multi-hop relaying can improve connectivity [18] as well as the coverage probability and the transmission capacity of a network [8]. However, a large number of hops requires significant signaling and scheduling overhead to determine the most suitable relays, which in turn can lead to increased energy consumption and latency. Also, the potential security issues associated with multi-hop relaying cannot be ignored [19], [20], as the points at which a malicious user can intercept the transmitted signal increase. On the other hand, 2-hop relaying in combination with relay beamforming can increase the communication range while avoiding most of the aforementioned issues [21], at the cost, however, of added complexity for the computation of beamforming weights and phase synchronization [22]. By additionally taking into consideration the 3D structure of the environment, signal blockage can also be avoided by optimal positioning of aerial mobile relays [23]. This approach has already been considered for supporting mmWave communication [24], in the context of Unmanned Aerial Vehicle (UAV) networks.

The most widely studied problem in 2-hop relaying for device-to-device outdoor communication is relay selection, i.e., determining the minimum number of deployed relays [25], and/or best relay configuration, in order to optimize the quality of communication [8], [26]–[29]. Typically, relay

selection requires instantaneous Channel State Information (CSI) between source and relays, and relays and destination. When CSI changes with time, optimal relay selection can become a resource-demanding task, as the network needs to *first* compute its channels via the exchange of pilots, and *then* decide on the best relay configuration. This process takes up bandwidth, wastes power and introduces latency in the network. In the following, any relay selection scheme that conforms to the aforementioned serial procedure will be referred to as *ideal*.

In this paper, we consider a relay-assisted beamforming approach for mmWave communications in an urban scenario, and propose a new, resource efficient relay selection scheme, designed to *optimally* enhance QoS in 2-hop Amplify-and-Forward (AF) cooperative networks. By exploiting the correlation structure of the channel induced by shadowing, a phenomenon prominent in mmWave communications [30], we are able to reduce both latency and CSI estimation overhead. AF is adopted herein as it is the simplest forwarding scheme, and of minimal implementation complexity. Other, possibly more involved forwarding schemes may also be considered; however, this is out of the scope of this work.

More specifically, our system model consists of static relays deployed in clusters across streets. Each cluster is defined within an area over which the channel exhibits similar statistical characteristics. This is typical in mmWave networks, which are primarily designed for relatively short distance point-to-point communications. Assuming a timeslotted system operation, the proposed scheme optimizes QoS in a 2-stage fashion, where, in every time slot, and simultaneously with AF beamforming to the destination, each cluster *predictively* selects a representative relay (1st stage) to optimally enhance AF beamforming at the subsequent time slot (2nd stage). Predictive relay selection is achieved by exploiting channel correlations with current and past networkwide magnitude-only CSI (also known as Received Signal Strength (RSS)) which is invariant to relay cluster size, and is measured sequentially during the operation of the system. This combination of cooperative beamforming and relay selection, which together comprise the proposed 2-stage problem formulation, presents distinct operational advantages over the trivial ideal scheme, stemming directly from the predictive nature of our approach.

In our setting, network QoS is quantified by the expected *Signal-to-Interference+Noise Ratio (SINR)* at the destination, which is a standard performance metric, and our goal is to maximize that expected SINR, subject to a shared power constraint among all relay clusters. Therefore, the optimal beamforming weights need to be computed centrally for all clusters. Nevertheless, the proposed relay selection procedure is conducted in a *completely distributed* manner; each cluster *independently* decides its successor representative for the subsequent time slot by solving a simple local stochastic optimization problem, *without* the need for inter-cluster information exchange.

Exploiting CSI correlations to predictively determine mo-



FIGURE 1. A birds-eye view of the assumed urban topology. Each numbered red circle depicts an intersection, that connects two or more road segments. A solid numbered blue line depicts a relay cluster placed along the respective street segment. The dotted lines depict all mmWave signal propagation paths from the source to the destination, passing through relay cluster 3.

bile relay movement in cooperative beamforming networks was recently explored in [22], in a conventional, free-space setting. Here, we use a similar idea for predictive relay selection. However, the special mmWave signal propagation characteristics make the treatment of the selection problem substantially different than that of [22]. In particular, the diversity resulting from the reflective nature of the mmWave signal propagation, as well as the possibly street-wise varying channel parameters induce significant differences as far as both problem description and development of corresponding efficient implementation techniques are concerned.

Along with our proposed adaptive relay selection scheme, this paper makes the following additional contributions:

1) We propose *distributed cooperative* beamforming for expected SINR maximization in mmWave networks. Our beamforming formulation alone allows for efficient exploitation of the spatial diversity induced by dominant mmWave propagation paths, which is a consequence of the spatial propagation patterns of the mmWave medium [30]–[32]. The CSI induced by the mmWave propagation paths is optimally combined constructively at the destination, resulting in superior network QoS, without the disadvantages of multi-hop relaying. Although distributed beamforming is a well established technique for exploitation of spatial diversity in free-space communications, this paper is the first to put forward its application to urban mmWave networks, which are inherently topologically distinct as compared to the free-space setting. Indeed, the spatial structure of the mmWave medium is explicitly reflected in the adopted mmWave channel model, as well as in the form of the optimal beamforming weights and achieved network SINR. In particular, our channel model extends the state of the art by introducing a new channel correlation kernel for effectively modeling the

statistical dependencies among the involved sourcerelay and relay-destination channels; such dependencies do not appear in free-space channel modeling.

- 2) As briefly mentioned above, within each time slot, the implementation of the proposed relay selection scheme is *completely decoupled* from that of optimal beamforming. This is due to the predictive nature of the proposed scheme, which determines the best cluster representative *before* the start of each time slot. Consequently, in a given slot, optimal beamforming and optimal predictive relay selection for the next slot can be performed in parallel, as one does not depend on the other. This parallelism results in improved time slot utilization. Additionally, as predictive selection is implemented solely based on past channel measurements, significant reduction of CSI estimation overhead per time slot is achieved as compared to the respective ideal scheme, with the reduction being more pronounced as the relay density per cluster increases. This is particularly important in mmWave networks, where dense infrastructure is essential for achieving satisfactory performance [33].
- 3) We propose a novel, practical and computationally efficient technique for implementing our proposed relay selection scheme. Specifically, the local stochastic problem each cluster is responsible for is replaced by a surrogate based on Sample Average Approximation (SAA) [34], which relies on *predictive Monte Carlo sampling* of the channel uncertainty involved. Heavily appealing to the statistical structure of the adopted mmWave channel model, the proposed technique efficiently exploits spatiotemporal correlations of the mmWave medium, and results in easily computable, near-optimal relay selection policies. This is achieved via a well-designed, non-trivial combination of Kalman filtering and Gaussian process regression.

The effectiveness of the proposed joint beamforming and relay selection system is confirmed through numerical simulations, conducted using synthetically generated CSI data. First, all our numerical results show that the SINR performance of our adaptive relay selection scheme is *almost identical* to the ultimate performance achieved via the respective ideal scheme. On the other extent, our simulations also verify that, as expected, the proposed strategic scheme clearly outperforms any randomized selection policy which does not exploit experience accumulated during system operation.

Additionally, we examine the effect of different cluster topologies on overall system performance. On the one hand, we confirm that expected network QoS increases with the number of clusters taking part in the communication, also revealing a probably sublinear relevant trend. On the other hand, our simulations corroborate that cluster placement in the city indeed affects the overall QoS of the network users, recognizing the importance of cluster assignment as an open problem for future research.

#### **II. MMWAVE URBAN CHANNEL MODEL**

This section is dedicated to the development of a sufficiently detailed urban mmWave channel model, which will be exploited throughout the rest of the paper. Our channel model can be applied to *any* city topology consisting of a densely built area with high-rise buildings, separated by non-curved street canyons [31], [35], [36].

To facilitate our presentation, we hereafter consider simplified city topologies such as that of Fig. 1, which shows a top view schematic of a particular urban area, where the numbered circles indicate road intersections, and the vertical and horizontal lines connecting those circles denote streets. Due to blockage caused by high-rise buildings, the only way a mmWave signal starting from a source located at  $\mathbf{p}_S$  can reach its destination at  $\mathbf{p}_D$  is by traversing street segments [31]. More specifically, the transmitted signal is spatially diversified through all sets of consecutive, nonrepeating segments from the source to the destination. Then, a (dominant) propagation path is defined as every such set of traversed street segments whose aggregate length is equal to the *minimum*  $\ell_1$ -distance from the source to the destination. We adopt the convention of [31], where the Lineof-Sight (LoS) portion of every path is the segment between the transmitting node and the nearest intersection, while the remaining segments comprise the Non-Line-of-Sight (NLoS) portion of the path. All considered paths have common LoS portion, while their NLoS portions differ.

To overcome severe signal attenuation, we deploy clusters of relays across certain street segments, which will beamform the signal to its destination. For simplicity in exposition, we assume that each cluster contains evenly spaced relays. We note, though, that our approach works for any spatial relay distribution within each cluster. At each time instance, only one relay from every cluster, namely, *the cluster representative*, is active. The relays are connected via fiber to a central node via which they can exchange information. A propagation path between the source or destination and any of the cluster representatives, as well as the corresponding LoS and NLoS portions of the path are all defined in exactly the same fashion as in the previous paragraph.

Let  $N_c$  be the number of available relay clusters in the network. Also, let  $L_r$  be the number of all possible signal paths from  $\mathbf{p}_S$  to relay cluster  $r = 1, \ldots, N_c$ . The channel between  $\mathbf{p}_S$  and a relay in cluster r located at  $\mathbf{p}$  is experienced as a combination of all channels across all possible paths between  $\mathbf{p}_S$  and  $\mathbf{p}$ . In particular, under the flat fading assumption, the complex channel gain from  $\mathbf{p}_S$  to point  $\mathbf{p}$  along path i can be decomposed as [37]

$$f_{ri}(\mathbf{p},t) \triangleq \underbrace{f_{ri}^{PL}(\mathbf{p})}_{path-loss \ shadowing \ multi-path} \underbrace{f_{ri}^{SH}(\mathbf{p},t)}_{multi-path} \underbrace{f_{ri}^{MF}(\mathbf{p},t)}_{multi-path}, \qquad (1)$$

where  $f_{ri}^{PL}(\mathbf{p})$  is the path-loss component,  $f_{ri}^{SH}(\mathbf{p},t)$  the large-scale fading component (shadowing), and  $f_{ri}^{MF}(\mathbf{p},t)$  the small-scale fading component (multi-path). A similar decomposition holds for the channel  $g_{ri}(\mathbf{p},t)$  from  $\mathbf{p}$  to

 $\mathbf{p}_D$ , along path  $i = 1, \dots, K_r$ , where  $K_r$  is the number of respective signal paths from cluster r to the destination.

In the mmWave setting, the channel path-loss does not depend on the Euclidean distance between  $\mathbf{p}_S$  and  $\mathbf{p}$ , but rather on their absolute locations (Manhattan distance), and is therefore parametrized separately for each segment [7], [31], [32], [35], [36], [38]. Let the set of all individually traversed street segments  $\tau$  of path *i* to cluster *r* be denoted by  $S_{ri}^f$ , which includes the segment  $\tau_S$  where the source is located, but does not include segment  $\tau_r$  where cluster *r* is located. Similarly, the set of traversed segments of path *i* from cluster *r* to the destination, including the segment  $\tau_D$  the destination is located but excluding segment  $\tau_r$ , is  $S_{ri}^g$ . In the following, we consider only the source-relay channels  $f_{ri}$ , for every path *i* associated with cluster *r*. The discussion for  $g_{ri}$  follows in a completely analog manner, and is omitted for brevity.

As in [31], we also assume an additional loss  $\Delta$  occurring at every intersection, i.e., every propagation path exhibits a total *intersection loss*  $\Delta N_r^f$ , where  $N_r^f$  are the number of traversed intersections from  $\mathbf{p}_S$  to  $\mathbf{p}$ . Therefore, the overall path-loss component of channel  $f_{ri}$  is expressed as

$$f_{ri}^{PL}(\mathbf{p}) = 10^{-\frac{\Delta N_r^f}{2 \cdot 10}} d_{\tau_S}^{-\frac{\alpha_L}{2}} (d_{\tau_r}^f(\mathbf{p}))^{-\frac{\alpha_N}{2}} \prod_{\tau \in \mathcal{S}_{ri}^f \setminus \{\tau_S\}} d_{\tau}^{-\frac{\alpha_N}{2}},$$
(2)

where  $d_{\tau_S}$  denotes the length of the LoS segment  $\tau_S$ ,  $d_{\tau_r}^f(\mathbf{p})$  is the distance between the intersection of segment  $\tau_r$  associated with  $f_{ri}$  and location  $\mathbf{p}$  in  $\tau_r$  (that is, the intersection of  $\tau_r$  which is  $\ell_1$ -closest to the source), and  $d_{\tau}$  denotes the length of the  $\tau$ -th street segment. We assume that a relay *cannot* be located *exactly on* an intersection, so  $d_{\tau_r}^f \neq 0$ .

Likewise, the shadowing and multi-path components of the channel experienced across each path may be decomposed on a per-segment basis as

$$f_{ri}^{SH}(\mathbf{p},t) = s_{\tau_r}^f(\mathbf{p},t) \prod_{\tau \in \mathcal{S}_{ri}^f} s_{\tau}(t) \quad \text{and} \tag{3}$$

$$f_{ri}^{MF}(\mathbf{p},t) = q_{\tau_r}^f(\mathbf{p},t) \prod_{\tau \in \mathcal{S}_{ri}^f} q_{\tau}(t),$$
(4)

where  $s_{\tau}$  and  $q_{\tau}$  are the shadowing and multi-path terms experienced across segment  $\tau$ .

Consequently, by expressing the magnitude of (1) in logarithmic scale (dB), we obtain the *additive model* 

$$F_{ri}\left(\mathbf{p},t\right) \triangleq 10\log_{10}\left(\left|f_{ri}^{PL}\left(\mathbf{p}\right) \cdot f_{ri}^{SH}\left(\mathbf{p},t\right) \cdot f_{ri}^{MF}\left(\mathbf{p},t\right)\right|^{2}\right)$$
$$\triangleq a_{ri}^{f}(\mathbf{p}) + b_{ri}^{f}(\mathbf{p},t) + c_{ri}^{f}(\mathbf{p},t)$$
(5)

where

$$-a_{ri}^{f}(\mathbf{p}) \triangleq \alpha_{L} 10 \log_{10} d_{\tau_{S}} + \alpha_{N} \sum_{\tau \in \mathcal{S}_{ri}^{f}} 10 \log_{10} d_{\tau}$$
$$+ a_{N} 10 \log_{10} d_{\tau_{r}}^{f}(\mathbf{p}) + \Delta N_{r}^{f}, \qquad (6)$$
$$b_{ri}^{f}(\mathbf{p}, t) \triangleq \sum_{\tau \in \mathcal{S}_{ri}^{f}} 10 \log_{10} |s_{\tau}(t)|^{2} + 10 \log_{10} |s_{\tau_{r}}^{f}(\mathbf{p}, t)|^{2}$$
$$\triangleq \sum_{\tau \in \mathcal{S}_{ri}^{f}} \beta_{\tau}(t) + \beta_{\tau_{r}}^{f}(\mathbf{p}, t) \quad \text{and} \qquad (7)$$

VOLUME 4, 2016

### **IEEE**Access

$$c_{ri}^{f}(\mathbf{p},t) \triangleq \sum_{\tau \in \mathcal{S}_{ri}^{f}} 10 \log_{10} |q_{\tau}(t)|^{2} + 10 \log_{10} |q_{\tau_{r}}^{f}(\mathbf{p},t)|^{2}$$
$$\triangleq \sum_{\tau \in \mathcal{S}_{ri}^{f}} \xi_{\tau}(t) + \xi_{\tau_{r}}^{f}(\mathbf{p},t).$$
(8)

We should note that in the above equations, the *combined* shadowing components pertaining to segments without relays have been separated from the respective term referring to the segment containing the relay cluster. Those terms exhibit distinct statistical behavior, and will be considered separately.

In addition to the above, for every time slot t, we assume a phase term  $e^{j2\pi\phi_{\tau}(t)}$  for each distinct segment  $\tau \in S_{ri}^{f}$ ,  $r = 1, \ldots, N_c, i = 1, \ldots, L_r$ , where each  $\phi_{\tau}(t)$  is uniformly distributed in [0,1]. Modeling the phase as a uniformly distributed process and independent of the respective channel magnitude is a standard assumption in the literature, which heuristically follows from the statistical structure of the wellknown Rayleigh fading model [37]. Additionally, as phase varies rapidly both in time and space, phase correlations are difficult to capture, therefore model, let alone exploit in statistical prediction. This fact consequently justifies our phase whiteness assumption, in both time and space. Note that phase whiteness is a standard assumption in the literature, and in various contexts (see, e.g., [37] [39]). Similarly, for every time slot t and every location  $\mathbf{p}$ , we assume another phase term  $\phi_{\tau_n}^f(\mathbf{p},t)$ , also uniformly distributed in [0, 1]. Across all time slots, all segments, and all locations, all phase components are mutually independent, and also independent of the respective channel magnitudes, as well. Then, the channel  $f_{ri}(\mathbf{p}, t)$  can be reconstructed as

$$f_{ri}(\mathbf{p},t) = e^{\frac{ln(10)F_{ri}(\mathbf{p},t)}{20}} e^{j2\pi\Phi_{ri}^{f}(\mathbf{p},t)},$$
(9)

where  $\Phi_{ri}^{f}(\mathbf{p}, t) \triangleq \sum_{\tau \in S_{ri}^{f}} \phi_{\tau}(t) + \phi_{\tau_{\tau}}^{f}(\mathbf{p}, t)$ . For all segments, we assume a log-normal distribution for

For all segments, we assume a log-normal distribution for modeling shadowing and multi-path fading [11]. The channel paths  $f_{ri}$ ,  $i = 1, ..., L_r$ , are statistically dependent, as they might traverse common segments. Still, it is reasonable to model all shadowing and multi-path components as being mutually independent across different segments, since each segment will exhibit distinct spatial features.

However, within each segment  $\tau \in S_{ri}^f$ ,  $\beta_{\tau}(t)$  is assumed to be zero mean and jointly Gaussian in time, with correlation between two time slots k and l given by

$$\mathbb{E}[\beta_{\tau}(k)\beta_{\tau}(l)] \triangleq \eta^2 e^{-|k-l|/\gamma}, \tag{10}$$

where  $\eta^2$  is the shadowing power and  $\gamma$  the correlation time [22]. Further assuming that the multi-path component  $q_{\tau}(t)$ is white in time with variance  $\sigma_{\xi}^2$  [39], the combined logmagnitude terms  $z_{\tau}(t) \triangleq \beta_{\tau}(t) + \xi_{\tau}(t), t = 1, \dots, N_T$  are jointly Gaussian with mean zero and covariance

$$\boldsymbol{\Sigma}_{\tau} \triangleq \eta^2 \begin{bmatrix} 1 & \dots & e^{-(N_T - 1)/\gamma} \\ \vdots & \ddots & \vdots \\ e^{-(N_T - 1)/\gamma} & \dots & 1 \end{bmatrix} + \sigma_{\xi}^2 \mathbf{I}_{N_T}$$

$$\triangleq \eta^2 \mathbf{T} + \sigma_{\xi}^2 \mathbf{I}_{N_T} \in \mathbb{R}^{N_T \times N_T}. \tag{11}$$

Likewise, the term  $\beta_{\tau_r}^f(\mathbf{p}, t)$ , corresponding to the segment where cluster r is located, is assumed to be jointly Gaussian, *both* in space and time. Specifically, we assume that the individual relays of cluster r can be located at a discrete set of  $\delta$  positions across the segment  $\tau_r$ . At two such positions, say  $\mathbf{p}_m$  and  $\mathbf{p}_n$ , and between any two time slots k and l, the spatiotemporal correlation of  $\beta_{\tau_r}(\mathbf{p}, t)$  is defined as [22]

$$\mathbb{E}[\beta_{\tau_r}^f(\mathbf{p}_n, k)\beta_{\tau_r}^f(\mathbf{p}_m, l)] \triangleq \mathsf{K}_{FF}(\mathbf{p}_n, \mathbf{p}_m)e^{-|k-l|/\gamma}, \quad (12)$$

where the spatial kernel  $K_{FF}$  is given by

$$\mathsf{K}_{FF}(\mathbf{p}_n, \mathbf{p}_m) \triangleq \eta^2 e^{-\|\mathbf{p}_n - \mathbf{p}_m\|_2/\beta}.$$
 (13)

We further assume that the "incoming" and "outgoing" shadowing terms  $\beta_{\tau_r}^f(\mathbf{p},t)$  and  $\beta_{\tau_r}^g(\mathbf{p},t)$  at positions  $\mathbf{p}_m$  and  $\mathbf{p}_n$ and between time slots k and l are *themselves* correlated as

$$\mathbb{E}[\beta_{\tau_r}^f(\mathbf{p}_n, k)\beta_{\tau_r}^g(\mathbf{p}_m, l)] \triangleq \mathsf{K}_{FG}(\mathbf{p}_n, \mathbf{p}_m)e^{-|k-l|/\gamma}, \quad (14)$$

where the *cross-correlation* kernel  $K_{FG}$  is defined as

$$\mathsf{K}_{FG}(\mathbf{p}_n, \mathbf{p}_m) \triangleq \eta^2 e^{(\epsilon \|\mathbf{p}_n - \mathbf{p}_m\|_2 - d_{max})/\beta}, \qquad (15)$$

with  $\epsilon = 1$  for  $d_{\tau_n}^f(\mathbf{p}_m) + d_{\tau_n}^g(\mathbf{p}_n) \ge d_{full}$  and  $\epsilon = -1$  otherwise, and where  $d_{full}$  is the length of segment  $\tau_r$ , and  $d_{max}$ is the furthest possible distance between two discrete relay positions of a cluster. This kernel describes the correlation between the incoming and outgoing channels at each cluster, at different locations and at different time slots. Intuitively, correlation should be proportional to the size of the part of the segment which is traversed by both channels, if such a part exists (see Fig. 2c). Otherwise, as the distance between the locations where the two channels are respectively experienced increases, their correlation should be decreasing (see Fig. 2b). The proposed kernel captures precisely the behavior outlined above, while resulting in a valid cross-covariance structure for  $\beta_{\tau_r}^f$  and  $\beta_{\tau_r}^g$ . As above, assuming that  $q_{\tau_r}^f(\mathbf{p},t)$ and  $q_{\tau_{\pi}}^{g}(\mathbf{p},t)$  are both white in *both* space and time, as well as mutually independent, the collection of combined terms

$$\begin{bmatrix} z_{\tau_r}^f(\mathbf{p}_i, t) \\ z_{\tau_r}^g(\mathbf{p}_i, t) \end{bmatrix} \triangleq \begin{bmatrix} \beta_{\tau_r}^f(\mathbf{p}_i, t) + \xi_{\tau_r}^f(\mathbf{p}_i, t) \\ \beta_{\tau_r}^g(\mathbf{p}_i, t) + \xi_{\tau_r}^g(\mathbf{p}_i, t) \end{bmatrix}, \quad (16)$$

for  $i = 1, ..., \delta$  and  $t = 1, ..., N_T$ , are Gaussian with mean zero and covariance  $\Sigma_{\tau_r} \in \mathbb{R}^{2\delta N_T \times 2\delta N_T}$  given by

$$\boldsymbol{\Sigma}_{\tau_r} \triangleq \mathbf{T} \otimes \mathbf{K} + \sigma_{\xi}^2 \mathbf{I}_{2\delta N_T}, \qquad (17)$$

where  $\otimes$  indicates Kronecker product, and the per-slot crosscovariance matrix  $\mathbf{K} \in \mathbb{R}^{2\delta \times 2\delta}$  is defined as

$$\mathbf{K} \triangleq \begin{bmatrix} \mathsf{K}_{FF} & \mathsf{K}_{FG} \\ \mathsf{K}_{FG} & \mathsf{K}_{GG} \end{bmatrix},\tag{18}$$

where, *overloading notation*,  $K_{FF}$ ,  $K_{GG}$  and  $K_{FG}$  are correlation matrices corresponding to the kernels (13) and (15), respectively, each evaluated on all  $\delta^2$  pairs of possible positions across segment  $\tau_r$ , according to some common order.



**FIGURE 2.** Cross-correlation structure of the incoming and outgoing channel terms  $\beta_{\tau_r}^f(\mathbf{p}_n, t), \beta_{\tau_r}^g(\mathbf{p}_m, t)$  when (a)  $d_{\tau_r}^f(\mathbf{p}_m) + d_{\tau_r}^g(\mathbf{p}_n) = d_{full}$ , (b)  $d_{\tau_r}^f(\mathbf{p}_m) + d_{\tau_r}^g(\mathbf{p}_n) < d_{full}$ , and (c)  $d_{\tau_r}^f(\mathbf{p}_m) + d_{\tau_r}^g(\mathbf{p}_n) > d_{full}$ , for cluster r with  $\delta = 3$  relay positions, at a common time slot t.

#### **III. JOINT BEAMFORMING AND RELAY SELECTION**

Determining the cluster topology, i.e., the number of deployed clusters as well as their locations is an important problem that will be studied in our future work, and could draw from current literature on optimal relay placement [25], [40], [41]. Our proposed scheme operates assuming that the clusters change at a low rate, and focuses on the time period over which the clusters have been optimally determined and are fixed. During that time, the statistical model of the channel stays the same; however, the channel itself changes. Therefore, assignment of new clusters is necessary only when "coarse-grained" features of the communication system change, such as the locations of the source and destination, the ergodic properties of the communication channel, and other similar statistics. Then, cluster assignment should be performed at a coarser time scale than the change of the individual relay channel realizations. For the duration of a communication task, the aforementioned features typically do not change rapidly, and therefore the cluster does not need to change, justifying the need for a relay selection scheme that operates at a per time-slot basis.

In every time slot, the proposed system *jointly* performs *beaforming* and *relay selection*, by addressing a 2-stage stochastic problem [22]. Before going into the details (and the advantages) of each stage separately, we should note that although the 2-stage problem refers to the necessary actions needed to be taken during a single time slot, in practice these actions refer to two consecutive time slots, due to the availability of the required CSI. More specifically, during time slot *t*, *both* current beamforming weights of the cluster representative are calculated (corresponding, as discussed

below, to the 2nd stage problem at time slot t), and the relays from all clusters to be selected at the next time slot are determined (which corresponds to the 1st stage problem at time slot t+1). Both tasks (beamforming and relay selection) are based on current CSI, as well as past CSI of cluster representatives selected up to time slot t.

We assume that 2-hop relaying is used to assist the communication between  $\mathbf{p}_S$  and  $\mathbf{p}_D$ . The whole network is assumed to operate for  $N_T$  time slots. In each time slot  $t = 1, \ldots, N_T$ , the source at  $\mathbf{p}_S$  transmits the signal  $\sqrt{P_S}s(t)$ , where s(t)is an information symbol with  $\mathbb{E}[|s(t)|^2] = 1$ , and  $P_S > 0$ the source transmission power. The signal received at the representative relay of *each* cluster r, located at  $\mathbf{p}_r(t)$  is,

$$R_r(t) = \sum_{i=1}^{L_r} \sqrt{P_S} f_{ri}(t) s(t) + n_r(t), \qquad (19)$$

where  $n_r(t) \sim C\mathcal{N}(0, \sigma^2)$  is the reception noise at cluster r. Working in an AF fashion, each cluster representative modulates its received signal  $R_r(t)$  by a complex weight  $w_r(t)$  and re-transmits it. Note that a mmWave signal arriving at  $\mathbf{p}_D$  directly from the source and without the help of a relay has negligible power, and can be ignored. Therefore, the aggregate signal received at the destination from all relay representatives is

$$y_{D}(t) = \sum_{r=1}^{N_{c}} \sum_{k=1}^{K_{r}} w_{r}(t) g_{rk}(t) R_{r}(t) + n_{D}(t)$$
  
=  $\underbrace{\sqrt{P_{S}} \sum_{r=1}^{N_{c}} w_{r}(t) \sum_{k=1}^{K_{r}} \sum_{i=1}^{L_{r}} g_{rk}(t) f_{ri}(t) s(t)}_{signal}}_{signal}$   
+  $\underbrace{\sum_{r=1}^{N_{c}} w_{r}(t) \sum_{k=1}^{K_{r}} g_{rk}(t) n_{r}(t) + n_{D}(t)}_{interference + destination noise}}$  (20)

where  $n_D(t) \sim C\mathcal{N}(0, \sigma_D^2)$  is the reception noise at  $\mathbf{p}_D$ .

#### A. OPTIMAL BEAMFORMING FOR 2-HOP RELAYING

We extend the distributed relay beamforming schemes studied in [42], [43], to the significantly more complex setting of urban mmWave relay networks. Here, distributed beamforming is considered for enforcing relay cluster cooperation, such that *all individual signal paths* forwarded from *all* relay clusters are combined constructively at the destination. Although the solution of the beamforming stage for the mmWave communication setting studied herein is a straightforward manipulation of the expressions found in the literature pertaining to the free-space communication scenario, [42], the result is interesting because it shows the explicit dependence of the optimal beamforming weights and achievable SINR on the *aggregate* mmWave channels from all propagation paths.

At every time slot t, the goal is to obtain the respective beamforming weights to be used by each cluster,  $\mathbf{w}(t) \triangleq$   $[w_1^*(t), \ldots, w_{N_c}^*(t)]^{\mathrm{T}} \in \mathbb{C}^{N_c \times 1}$ , such that the SINR at  $\mathbf{p}_D$  is maximized, subject to a total transmission power budget  $P_C > 0$  over all relay clusters. Define the vectors

$$\mathbf{f}_{r}(\mathbf{p},t) \triangleq [f_{r1}(\mathbf{p},t),\ldots,f_{rL_{r}}(\mathbf{p},t)]^{\mathrm{T}} \in \mathbb{C}^{L_{r} \times 1}, \quad (21)$$

$$\mathbf{g}_{r}(\mathbf{p},t) \triangleq [g_{r1}(\mathbf{p},t),\ldots,g_{rK_{r}}(\mathbf{p},t)]^{\mathrm{T}} \in \mathbb{C}^{K_{r} \times 1}, \quad (22)$$

 $r = 1, \ldots, N_c$ . Then, after dropping dependence on (t) and  $(\mathbf{p}_r(t), t)$  for brevity, the SINR is maximized by solving [42]

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \frac{\mathbf{w}^{\text{H}} \mathbf{R} \mathbf{w}}{\mathbf{w}^{\text{H}} \mathbf{Q} \mathbf{w} + \sigma_D^2}, \\ \text{subject to} & \mathbf{w}^{\text{H}} \mathbf{D} \mathbf{w} \leq P_C \end{array}$$
(23)

where

$$\mathbf{R} \triangleq P_S \mathbf{h} \mathbf{h}^{\mathrm{H}}, \ \mathbf{h} \triangleq [\mathbf{1}^{\mathrm{T}} \mathbf{g}_1 \mathbf{1}^{\mathrm{T}} \mathbf{f}_1, \dots, \mathbf{1}^{\mathrm{T}} \mathbf{g}_{N_c} \mathbf{1}^{\mathrm{T}} \mathbf{f}_{N_c}]^{\mathrm{T}}, \quad (24)$$

$$\mathbf{D} \triangleq P_S \operatorname{diag}(|\mathbf{1}^{\mathrm{T}} \mathbf{f}_1|^2, \dots, |\mathbf{1}^{\mathrm{T}} \mathbf{f}_{N_c}|^2) + \sigma^2 \mathbf{I}_{N_c} \text{ and } (25)$$

$$\mathbf{Q} \triangleq \sigma^2 \operatorname{diag}\left(|\mathbf{1}^{\mathrm{T}}\mathbf{g}_1|^2, \dots, |\mathbf{1}^{\mathrm{T}}\mathbf{g}_{N_c}|^2\right)$$
(26)

and 1 is the all-ones vector. A crucial technical property of problem (23) is that its optimal can be explicitly expressed as [22], [42]

$$V(t) = \sum_{r=1}^{N_c} \frac{P_C P_S |\mathbf{1}^{\mathrm{T}} \mathbf{f}_r|^2 |\mathbf{1}^{\mathrm{T}} \mathbf{g}_r|^2}{P_S \sigma_D^2 |\mathbf{1}^{\mathrm{T}} \mathbf{f}_r|^2 + P_C \sigma^2 |\mathbf{1}^{\mathrm{T}} \mathbf{g}_r|^2 + \sigma^2 \sigma_D^2}$$
  
=  $\sum_{r=1}^{N_c} V_I (\mathbf{S}_r(\mathbf{p}_r(t), t)),$  (27)

where  $S_r(\mathbf{p}_r(t), t)$  is a vector of all random variables referring to the shadowing, multi-path, and phase terms of all *unique* segments traversed for all paths from  $\mathbf{p}_S$  to  $\mathbf{p}_D$ , which also pass through each cluster representative, located at  $\mathbf{p}_r(t)$ . As can be seen from (27), V(t) depends on the relay positions at time slot t. Thus, by optimally positioning the relays, V(t) can be further maximized. This problem is explored in the next subsection.

Interestingly enough, it turns out that the optimal beamforming vector that achieves (27) enjoys an explicit form similar to that used in the free-space scenario [42], i.e.,

$$\mathbf{w}_{opt}(t) = \sqrt{P_C} \mathbf{D}^{-\frac{1}{2}} \frac{\mathbf{v}_{max}}{\|\mathbf{v}_{max}\|_2},$$
(28)

where the *alignment vector*  $\mathbf{v}_{max} \in \mathbb{C}^{N_c \times 1}$  is defined as

$$\mathbf{v}_{max} \triangleq \begin{bmatrix} \frac{P_{S}\mathbf{1}^{\mathrm{T}}\mathbf{g}_{1}^{*}\mathbf{1}^{\mathrm{T}}\mathbf{f}_{1}^{*}}{P_{S}\sigma_{D}^{2}|\mathbf{1}^{\mathrm{T}}\mathbf{f}_{1}|^{2} + P_{C}\sigma^{2}|\mathbf{1}^{\mathrm{T}}\mathbf{g}_{1}|^{2} + \sigma^{2}\sigma_{D}^{2}} \\ \vdots \\ \frac{P_{S}\mathbf{1}^{\mathrm{T}}\mathbf{g}_{N_{c}}^{*}\mathbf{1}^{\mathrm{T}}\mathbf{f}_{N_{c}}^{*}}{P_{S}\sigma_{D}^{2}|\mathbf{1}^{\mathrm{T}}\mathbf{f}_{N_{c}}|^{2} + P_{C}\sigma^{2}|\mathbf{1}^{\mathrm{T}}\mathbf{g}_{N_{c}}|^{2} + \sigma^{2}\sigma_{D}^{2}} \end{bmatrix}, (29)$$

and where  $|\mathbf{1}^{T}\mathbf{f}_{r}|^{2}$  and  $|\mathbf{1}^{T}\mathbf{g}_{r}|^{2}$  are the incoming and outgoing aggregate channels at  $\mathbf{p}_{r}$ , respectively. In other words, each cluster representative at  $\mathbf{p}_{r}$  does *not* need to estimate the individual channels from every propagation path, but rather only the aggregate channel from all propagation paths. In practice, this can be computed by the selected relay via the exchange of pilots. One may also observe that, while the i-th element of  $\mathbf{v}_{max}$  can be estimated by the *i*-th relay only, the vector norm in (28) involves the source and destination channels of all cluster representatives who will beamform at the current time. Therefore, that scalar will have to be computed centrally and then distributed to all clusters. This can be done through a high-speed, optical fiber based, backhaul network, that connects all relay clusters, as well as all relays within a cluster, with each other [33], [44]. Clearly, for the source and destination which are, e.g. moving vehicles, no wired connection to the backhaul exists. The beamforming stage requires  $\mathcal{O}(N_c)$  operations.

As a final remark, we should note that, in practice, during this beamforming step, phase synchronization is required to take care of local oscillator phase offsets. Distributed beamforming synchronization is an active field of research [45], [46], that has also been studied in the context of mmWaves [47]–[49]. Here, we assume that the backhaul network can take care of the synchronization between the relay clusters.

#### B. OPTIMAL RELAY SELECTION FOR 2-HOP RELAYING

At every time slot, each cluster must decide which is the appropriate relay to be used for beamforming. Typically, this would *first* require estimating the respective channel of every relay in the cluster, and *then* deciding upon the strongest one. Clearly, this decision making procedure not only wastes power and bandwidth during CSI estimation, but also induces extra delay *before* optimized communication can take place *within* each time slot. This delay is significant, especially if the number of relays per cluster is large. In this subsection, we propose a new scheme for adaptive relay selection *which completely avoids this overhead*, thus resulting in much better time slot utilization.

More specifically, the proposed relay selection scheme is based on *transferring* the implementation of the relay selection procedure from the current time slot, to the previous time slot. In other words, relay selection would be implemented *predictively* by efficiently exploiting the statistical model of the mmWave channel, before the respective time slot starts. As a result, at each time slot, optimal beamforming is implemented by utilizing the cluster representatives which were optimally selected during the previous slot. This immediately results in the complete elimination of the "waiting delay" discussed above; indeed, if predictive relay selection is sufficiently accurate, then the cluster representatives at each time slot can be *predetermined*, *before* the slot starts. This means that relay selection and beamforming can be completely decoupled within each time slot, and thus can be parallelized; indeed, at each time slot, optimal beamforming can be implemented simultaneously with optimal predictive relay selection affecting the next time slot. In addition to eliminating the "waiting delay", the proposed scheme also enables a substantial reduction of the CSI estimation overhead required for relay selection, as well as significant power savings. See Section IV for a more detailed discussion.

We now describe the proposed relay selection scheme in

detail. As described above, at time slot t, we are interested in deciding on the best relay representatives from *all* clusters to participate in beamforming at time slot t + 1, such that the networkwide SINR, V(t + 1), is maximized. However, at the current time slot t, future CSI needed for evaluating V(t + 1) is not yet available. Nevertheless, a reasonable causal criterion for optimal relay selection is to maximize a projection of V(t + 1) on information available at time slot t. Following this path, we propose to maximize an *MMSE* predictor of V(t + 1) relative to the collection  $C_r(t)$  of all magnitude CSI, or RSS, from the segments of all propagation paths associated with all previously selected representatives of cluster r, as well as the positions of the representatives themselves, up until and including t. Then, due to the additive structure of (27), each cluster r can independently solve

$$\begin{array}{ll} \underset{\mathbf{p}}{\text{maximize}} & \mathbb{E}\left[V_{I}(\boldsymbol{S}_{r}(\mathbf{p},t+1))|\mathcal{C}_{r}(t)\right]\\ \text{subject to} & \mathbf{p} \in \mathscr{C}_{r}(t) \end{array}, \quad (30)$$

where  $\mathscr{C}_r(\cdot)$  constitutes the set of candidate relays within the cluster which can potentially be selected. This set can either be *unconstrained*, including any relay within the cluster, or *constrained* to only a subset of relays within the cluster. As can be seen by (30), our approach exploits spatial and temporal dependencies of channel shadowing, an otherwise negative effect imposed by the communication medium, in order to actually benefit network QoS by predicting future, one-step-ahead SINR.

Next, define the sets  $S_r^f = \bigcup_{i=1}^{L_r} S_{ri}^f$  and  $S_r^g = \bigcup_{i=1}^{K_r} S_{ri}^g$ . Then, at every feasible location  $\mathbf{p} \in \mathscr{C}_r(t)$ , the objective of (30) may be expressed as

$$\mathbb{E}[V_I(\boldsymbol{S}_r(\mathbf{p},t+1)) | \mathcal{C}_r(t)] \\ = \int V_I(\mathbf{p},\boldsymbol{s}) p_{\boldsymbol{S}_r(\mathbf{p},t+1) | \mathcal{C}_r(t)}(\boldsymbol{s}) \mathrm{d}\boldsymbol{s},$$
(31)

where, dropping dependence on  $(\mathbf{p}, t+1)$ ,  $V_I$  may be reexpressed in a more integration-friendly form as

$$V_{I}(\cdot, \mathcal{Z}_{r}^{f}, \varphi_{r}^{f}, \mathcal{Z}_{r}^{g}, \varphi_{r}^{g}) = \frac{P_{C}P_{S}\mathsf{F}(\mathcal{Z}_{r}^{f}, \varphi_{r}^{f})\mathsf{G}(\mathcal{Z}_{r}^{g}, \varphi_{r}^{g})}{P_{S}\sigma_{D}^{2}\mathsf{F}(\mathcal{Z}_{r}^{f}, \varphi_{r}^{f}) + P_{C}\sigma^{2}\mathsf{G}(\mathcal{Z}_{r}^{g}, \varphi_{r}^{g}) + \sigma^{2}\sigma_{D}^{2}},$$
(32)

with F being a function of the sets  $Z_r^f = \{z_{\tau_r}^f, \{z_\tau\}_{\tau \in S_r^f}\}$ , and  $\varphi_r^f = \{\phi_{\tau_r}^f, \{\phi_\tau\}_{\tau \in S_r^f}\}$ , corresponding to the combined shadowing plus multi-path, and phase terms of the unique segments traversed in all paths between the source and cluster r, and respectively for G,  $Z_r^g$  and  $\varphi_r^g$ . Analytical expressions for F and G are presented in Appendix A.

By a slightly tedious but straightforward procedure, it may be shown that the joint conditional density of all random variables contained in vector  $S_r(\mathbf{p}, t+1)$  relative to  $C_r(t)$ can be expressed as (by overloading notation)

$$p_{\mathbf{S}_{r}(\mathbf{p},t+1)|\mathcal{C}_{r}(t)}(\mathcal{Z}_{r}^{f},\varphi_{r}^{f},\mathcal{Z}_{r}^{g},\varphi_{r}^{g})$$

$$= \mathcal{N}([z_{\tau_{r}}^{f}z_{\tau_{r}}^{g}];\boldsymbol{\mu}_{\tau_{r}}^{t+1|t}(\mathbf{p}),\boldsymbol{\Sigma}_{\tau_{r}}^{t+1|t}(\mathbf{p}))$$

$$\times \mathcal{U}(\phi_{\tau_{r}}^{f};0,1)\mathcal{U}(\phi_{\tau_{r}}^{g};0,1)$$

$$\times \prod_{\tau \in \mathcal{S}_r^f \cup \mathcal{S}_r^g} \mathcal{N}(z_\tau; \mu_\tau^{t+1|t}, (\sigma_\tau^{t+1|t})^2) \mathcal{U}(\phi_\tau; 0, 1),$$
(33)

where  $\mathcal{U}(\cdot; 0, 1)$  denotes the uniform density on [0, 1], and where  $\mu_{\tau_r}^{t+1|t}(\mathbf{p})$ ,  $\Sigma_{\tau_r}^{t+1|t}(\mathbf{p})$ ,  $\mu_{\tau}^{t+1|t}$  and  $(\sigma_{\tau}^{t+1|t})^2$  constitute the corresponding posterior statistics, each relative to *local* CSI at the corresponding segment, respectively. This readily follows by mutual independence of the corresponding CSI processes across segments. The derivation of (33) can be found in Appendix B.

Note that, although phase information at time slot t + 1 is present in (32), the objective (31) is independent of phase information *at past time slots*. This is because of the standard assumption that, for each segment, the phase component of the channel is white in time and space, and mutually independent of the respective phase component of all other segments. Indeed, one may readily observe that, in (33), all distributions associated with channel phases are uniform in [0, 1], which is precisely the prior distribution of all phase components, for all segments taking part in the communication.

From the discussion above, it follows that tractably evaluating (31) is a challenging task. Of course, as expected, the first step towards evaluation of (31) is the efficient determination of the aforementioned predictors. This is the subject of the next two subsections.

#### 1) Channel Prediction for Cluster-free Segments

The shadowing component of the channel for a cluster-free segment  $\tau$ ,  $\beta_{\tau}$ , is a Gaussian process evolving in time, which may also be represented as a *stable autoregression of order* 1. Indeed, it may be easily shown that, at every segment  $\tau \in \bigcup_{r=1}^{N_c} (S_r^f \cup S_r^g), \beta_{\tau}$  can be represented via the stochastic difference equation [50]

$$\beta_{\tau}(t) = \kappa \beta_{\tau}(t-1) + w_{\tau}(t), \quad t = 1, \dots, N_T, \quad (34)$$

where  $\kappa \triangleq e^{-1/\gamma}$ ,  $\beta_{\tau}(0) \sim \mathcal{N}(0, \eta^2)$ , with the latter being independent of  $w_{\tau}(t) \stackrel{i.i.d.}{\sim} \mathcal{N}(0, (1-\kappa^2)\eta^2)$ ,  $t = 1, \ldots, N_T$ .

At the same time, across time slots, the shadowing process  $\beta_{\tau}(t)$  cannot be measured directly. Instead, it may be considered as *corrupted by unpredictable noise*, due to the presence of the multi-path component  $\xi_{\tau}(t)$ ; indeed, at each time slot t and segment  $\tau$ , the term  $z_{\tau}(t) = \beta_{\tau}(t) + \xi_{\tau}(t)$  is observed.

Now, for every segment  $\tau \in \bigcup_{r=1}^{N_c} (\mathcal{S}_r^f \cup \mathcal{S}_r^g)$ , define the vector

$$\mathbf{m}_{\tau}^{1:t} \triangleq [z_{\tau}(1), \dots, z_{\tau}(t)]^{\mathrm{T}} \in \mathbb{R}^{t \times 1},$$
(35)

which contains all observable zero-mean CSI magnitudes associated with that segment, up to time t. Then, exploiting the autoregressive representation of (34), it follows that the posterior distribution of  $z_{\tau}(t+1)$  relative to  $\mathbf{m}_{\tau}^{1:t}$  is Gaussian with conditional mean and variance given by

$$\mu_{\tau}^{t+1|t} = \kappa \beta_{\tau}^{t|t} \quad \text{and} \tag{36}$$

$$(\sigma_{\tau}^{t+1|t})^2 = \kappa^2 \rho_{\beta_{\tau}}^{t|t} + (1-\kappa^2)\eta^2 + \sigma_{\xi}^2, \qquad (37)$$

respectively, where, by definition,

$$\beta_{\tau}^{t|t} \triangleq \mathbb{E}[\beta_{\tau}(t)|\mathbf{m}_{\tau}^{1:t}] \quad \text{and}$$
 (38)

$$\rho_{\beta_{\tau}}^{t|t} \triangleq \mathbb{E}[(\beta_{\tau}(t) - \mathbb{E}[\beta_{\tau}(t)|\mathbf{m}_{\tau}^{1:t}])^2 |\mathbf{m}_{\tau}^{1:t}]$$
(39)

are the conditional mean and variance of  $\beta_{\tau}(t)$  relative to  $\mathbf{m}_{\tau}^{1:t}$ , respectively. Therefore, determination of  $\mu_{\tau}^{t+1|t}$  and  $(\sigma_{\tau}^{t+1|t})^2$  is equivalent to that of  $\beta_{\tau}^{t|t}$  and  $\rho_{\beta_{\tau}}^{t|t}$ , respectively, for all  $t = 1, \ldots, N_T$ . Again due to the autoregressive structure of (34), the latter pair of conditional estimates may be evaluated *recursively* via a Kalman filter, achieving constant computational complexity per time slot. Specifically, for every  $t = 1, \ldots, N_T$ , both  $\beta_{\tau}^{t|t}$  and  $\rho_{\beta_{\tau}}^{t|t}$  may be evaluated recursively via the updates

$$\beta_{\tau}^{t|t} = \kappa \beta_{\tau}^{t-1|t-1} + K^{t}(z_{\tau}(t) - \kappa \beta_{\tau}^{t-1|t-1}), \qquad (40)$$

$$\rho_{\beta_{\tau}}^{t|t} = K^t \sigma_{\xi}^2 \quad \text{and} \tag{41}$$

$$K^{t} = \frac{\kappa^{2} \rho_{\beta_{\tau}}^{t-1|t-1} + (1-\kappa^{2})\eta^{2}}{\kappa^{2} \rho_{\beta_{\tau}}^{t-1|t-1} + (1-\kappa^{2})\eta^{2} + \sigma_{\xi}^{2}},$$
(42)

initialized by setting  $\beta_{\tau}^{0|0} = 0$  and  $\rho_{\beta_{\tau}}^{0|0} = \eta^2$ , stemming from the statistics of the initial condition  $\beta_{\tau}(0)$ . By direct comparison of (36) and (37) to the Kalman filter equations (40), (41) and (42), it is easy to derive an algorithm for the *direct* recursive evaluation of  $\mu_{\tau}^{t+1|t}$  and  $(\sigma_{\tau}^{t+1|t})^2$ , comprised, for  $t = 1, \ldots, N_T$ , by the dynamic equations

$$\mu_{\tau}^{t+1|t} = \kappa (1 - K^t) \mu_{\tau}^{t|t-1} + \kappa K^t z_{\tau}(t), \tag{43}$$

$$(\sigma_{\tau}^{t+1|t})^2 = (1+\kappa^2 K^t)\sigma_{\xi}^2 + (1-\kappa^2)\eta^2$$
 and (44)

$$K^{t} = \frac{(\sigma_{\tau}^{t|t-1})^{2} - \sigma_{\xi}^{2}}{(\sigma_{\tau}^{t|t-1})^{2}},$$
(45)

initialized by setting  $\mu_{\tau}^{1|0} = 0$  and  $(\sigma_{\tau}^{1|0})^2 = \eta^2 + \sigma_{\xi}^2$ . For each cluster-free segment  $\tau \in \bigcup_{r=1}^{N_c} (S_r^f \cup S_r^g)$ , the

For each cluster-free segment  $\tau \in \bigcup_{r=1}^{r} (S_r^f \cup S_r^g)$ , the corresponding Kalman filter may be implemented either centrally within each cluster, or in a completely distributed fashion, where each cluster-free segment is responsible for tracking its own channel, and then for distributing its estimate to the associated cluster, responsible for the actual relay selection. Due to its recursive nature, each Kalman filter required an order of  $\mathcal{O}(1)$  operations for each cluster-free segment.

#### Channel Prediction for Segments Containing Clusters

Next, consider the segment  $\tau_r$ , containing cluster r. Then, if we define  $\mathbf{z}_{\tau_r}^{f,g} \triangleq [z_{\tau_r}^f, z_{\tau_r}^g]^T$  and store all CSI measurements of every previously selected representative of cluster r in

$$\mathbf{m}_{\tau_r}^{1:t} = [\boldsymbol{z}_{\tau_r}^{f,g}(\mathbf{p}_r(1), 1), \dots, \boldsymbol{z}_{\tau_r}^{f,g}(\mathbf{p}_r(t), t)]^{\mathrm{T}} \in \mathbb{R}^{2t \times 1},$$
(46)

then, for each location  $\mathbf{p} \in \mathscr{C}_r(t)$ , the mean vector and covariance matrix of the Gaussian random vector  $\mathbf{z}_{\tau_r}^{f,g}(\mathbf{p},t+1)$ conditioned on  $\mathbf{m}_{\tau_r}^{1:t}$  are

$$\boldsymbol{\mu}_{\tau_r}^{t+1|t}(\mathbf{p}) = (\bar{\boldsymbol{\sigma}}_{\tau_r}^{1:t}(\mathbf{p}))^{\mathrm{T}} (\bar{\boldsymbol{\Sigma}}_{\tau_r}^{1:t})^{-1} \mathbf{m}_{\tau_r}^{1:t} \in \mathbb{R}^{2 \times 1}$$
(47)

$$\boldsymbol{\Sigma}_{\tau_r}^{t+1|t}(\mathbf{p}) = \bar{\mathbf{K}} - (\bar{\boldsymbol{\sigma}}_{\tau_r}^{1:t}(\mathbf{p}))^{\mathrm{T}} (\bar{\boldsymbol{\Sigma}}_{\tau_r}^{1:t})^{-1} \bar{\boldsymbol{\sigma}}_{\tau_r}^{1:t}(\mathbf{p}) \in \mathbb{R}^{2 \times 2},$$
(48)

respectively, where

$$\bar{\mathbf{K}} = \begin{bmatrix} \eta^2 + \sigma_{\xi}^2 & \eta^2 e^{-d_{max}/\beta} \\ \eta^2 e^{-d_{max}/\beta} & \eta^2 + \sigma_{\xi}^2 \end{bmatrix},$$
(49)

and  $\bar{\Sigma}_{\tau_r}^{1:t} \in \mathbb{R}^{2t \times 2t}$ ,  $\bar{\sigma}_{\tau_r}^{1:t} \in \mathbb{R}^{2t \times 2}$  are sampled for every time slot until t from  $\Sigma_{\tau_r} \in \mathbb{R}^{2\delta N_T \times 2\delta N_T}$ , at the positions that correspond to the distance between the candidate location **p** and the respective locations where the incoming and outgoing channels of segment  $\tau_r$  have been experienced so far. We should note that unlike before, (47) and (48) cannot be estimated using a Kalman filter, but rather, using full-blown Gaussian process regression. The dominant operation of (47) and (48) is the inversion of the covariance matrix  $\bar{\Sigma}_{\tau_r}^{1:t}$ . The computational complexity of this inversion is of the order of  $\mathcal{O}(t^3)$  operations, and grows with time due to conditioning on past CSI. Nevertheless, the complexity can be reduced to  $\mathcal{O}(t^2)$ , via a classical application of the matrix inversion lemma; for details, see ([22], section VI).

#### 3) Reduced-Complexity Sample Average Approximation

Having determined the necessary posterior statistics involved in (33), our next step would be to evaluate the objective of (30) or, equivalently, the multi-dimensional integral (31). However, to the best of our knowledge, a closed-form representation of (31) is impossible to derive. Therefore, we resort to a near-optimal approach. In particular, we rely on the SAA method, and replace (30) by an easily computable *surrogate*, constructed via *unconditional* Monte Carlo sampling.

To define the proposed surrogate to (30), fix  $\mathbf{p} \in \mathscr{C}_r(t)$  and  $t = 1, \ldots, N_T$ , and consider the change of variables (again, overloading notation)

$$v_{\tau} = (\sigma_{\tau}^{t+1|t})^{-1} (z_{\tau} - \mu_{\tau}^{t+1|t}), \ \forall \tau \in \mathcal{S}_r^f \cup \mathcal{S}_r^g \text{ and } (50)$$

$$\boldsymbol{v}_{\tau_r}^{f,g} = \left(\boldsymbol{\Sigma}_{\tau_r}^{t+1|t}(\mathbf{p})\right)^{-1/2} \left(\boldsymbol{z}_{\tau_r}^{f,g} - \boldsymbol{\mu}_{\tau_r}^{t+1|t}(\mathbf{p})\right),\tag{51}$$

to the integral of (31). Additionally, also define the collections  $\mathcal{V}_r^f \triangleq \{ \boldsymbol{v}_{\tau_r}^{f,g}, \{v_{\tau}\}_{\tau \in \mathcal{S}_r^f} \}$  and  $\mathcal{V}_r^g \triangleq \{ \boldsymbol{v}_{\tau_r}^{f,g}, \{v_{\tau}\}_{\tau \in \mathcal{S}_r^g} \}$ . Then, (31) may be equivalently represented as

$$\mathbb{E}[V_I(\boldsymbol{S}_r(\mathbf{p},t+1))|\mathcal{C}_r(t)] = \int \bar{V}_I^{t+1|t}(\mathbf{p},\boldsymbol{s})p_{\bar{\boldsymbol{S}}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s},$$
(52)

where

$$\bar{V}_{I}^{t+1|t}(\mathbf{p}, \mathcal{V}_{r}^{f}, \varphi_{r}^{f}, \mathcal{V}_{r}^{g}, \varphi_{r}^{g})$$

$$= \frac{P_{C}P_{S}\mathsf{F}^{t+1|t}(\mathbf{p}, \mathcal{V}_{r}^{f}, \varphi_{r}^{f})\mathsf{G}^{t+1|t}(\mathbf{p}, \mathcal{V}_{r}^{g}, \varphi_{r}^{g})}{P_{S}\sigma_{D}^{2}\mathsf{F}^{t+1|t}(\mathbf{p}, \mathcal{V}_{r}^{f}, \varphi_{r}^{f}) + P_{C}\sigma^{2}\mathsf{G}^{t+1|t}(\mathbf{p}, \mathcal{V}_{r}^{g}, \varphi_{r}^{g}) + \sigma^{2}\sigma_{D}^{2}}$$
(53)

with the functions  $F^{t+1|t}$  and  $G^{t+1|t}$  being defined as<sup>1</sup>

$$\mathsf{F}^{t+1|t} (\mathbf{p}, \mathcal{V}_{r}^{f}, \varphi_{r}^{f}) \\ \triangleq \mathsf{F} \Big( \mathbf{p}, \big( \mathbf{\Sigma}_{\tau_{r}}^{t+1|t} (\mathbf{p}) \big)^{1/2} \boldsymbol{v}_{\tau_{r}}^{f,g} + \boldsymbol{\mu}_{\tau_{r}}^{t+1|t} (\mathbf{p}) \big|_{1}, \\ \{ \boldsymbol{v}_{\tau} \boldsymbol{\sigma}_{\tau}^{t+1|t} + \boldsymbol{\mu}_{\tau}^{t+1|t} \}_{\tau \in \mathcal{S}_{r}^{f}}, \varphi_{r}^{f} \Big)$$
(54)

<sup>1</sup>If  $\boldsymbol{x}$  is a vector,  $\boldsymbol{x}|_i$  denotes its *i*-th entry

and

$$\begin{aligned} \mathsf{G}^{t+1|t} \big(\mathbf{p}, \mathcal{V}_{r}^{g}, \varphi_{r}^{g}\big) \\ &\triangleq \mathsf{G} \Big(\mathbf{p}, \big(\mathbf{\Sigma}_{\tau_{r}}^{t+1|t}(\mathbf{p})\big)^{1/2} \boldsymbol{v}_{\tau_{r}}^{f,g} + \boldsymbol{\mu}_{\tau_{r}}^{t+1|t}(\mathbf{p})\big|_{2}, \\ &\{v_{\tau}\sigma_{\tau}^{t+1|t} + \boldsymbol{\mu}_{\tau}^{t+1|t}\}_{\tau \in \mathcal{S}_{r}^{g}}, \varphi_{r}^{g}\Big), \end{aligned} \tag{55}$$

and where  $ar{m{S}}$  follows the distribution induced by the density

$$p_{\bar{\boldsymbol{s}}}(\mathcal{V}_{r}^{f},\varphi_{r}^{f},\mathcal{V}_{r}^{g},\varphi_{r}^{g}) = \mathcal{N}(\boldsymbol{v}_{\tau_{r}}^{f,g};\boldsymbol{0},\mathbf{I}_{2})\mathcal{U}(\phi_{\tau_{r}}^{f};0,1)\mathcal{U}(\phi_{\tau_{r}}^{g};0,1) \times \prod_{\tau \in \mathcal{S}_{r}^{f} \cup \mathcal{S}_{r}^{g}} \mathcal{N}(z_{\tau};0,1)\mathcal{U}(\phi_{\tau};0,1).$$
(56)

The representation (52) exhibits an important and rather practically appealing property: The density  $p_{\bar{S}}$  is *completely independent of both* **p** and  $C_r(t)$ , and all such dependence has been transferred to  $\bar{V}_I^{t+1|t}$ . Consequently, *sampling* from  $p_{\bar{S}}$  is greatly facilitated, and this fact is exactly what makes our proposed SAA-based scheme attractive, whose description now follows.

Algorithm 1	l Ioint	Beam	form	ing	& I	Relav	Selectio	n Scheme
angor tunni .	L JOINT	Deam	uom	ung		i conu y	Derectio	in Seneme

e ;
for $t = 1 : N_T$ do
<b>Beamforming</b> (2nd stage of time slot $t$ )
<b>Inputs:</b> Channel aggregates $ 1^{T}\mathbf{f}_{r} ^{2}$ and $ 1^{T}\mathbf{g}_{r} ^{2}$ .
Compute optimal $\mathbf{w}_{opt}(t)$ from (28)
<b>Relay selection</b> (1st stage of time slot $t + 1$ )
for every cluster r do
<b>Inputs:</b> a) RSS $\{z_{\tau}(t)\}_{\tau \in S^{f}_{\tau} \cup S^{g}_{\tau}}$ until t.
b) RSS $z_{\tau_r}^f(t)$ and $z_{\tau_r}^g(t)$ until t.
Generate $\bar{S}$ from (56)
for each $\mathbf{p}\in \mathscr{C}_r(t)$ do
Compute $\hat{V}_I(\mathbf{p}, t+1)$ from (57)
end for
Choose $\mathbf{p}_r(t+1) \in \operatorname{argmax}_{\mathbf{p} \in \mathscr{C}_r(t)} \hat{V}_I(\mathbf{p}, t+1)$
end for
end for

For each relay cluster r and at every time slot t, the SAA method works by randomly generating a total of  $N_S$  scenarios, drawn from the distribution induced by  $p_{\bar{S}}$ . Clearly, due to the special form of  $p_{\bar{S}}$ , this is straightforward to implement. Then, each scenario  $\bar{S}^{(i)}$ ,  $i = 1, \ldots, N_S$  is used to evaluate  $\bar{V}_I^{t+1|t}$ , at every possible relay position within the set of feasible locations,  $\mathscr{C}_r(t)$ . Finally, leveraging (52), the SAA of (31) is formulated by replacing the expectation in its objective with an *empirical mean* as

$$\begin{array}{ll} \underset{\mathbf{p}}{\text{maximize}} & \hat{V}_{I}(\mathbf{p}, t+1) \triangleq \frac{1}{N_{S}} \sum_{i=1}^{N_{S}} \bar{V}_{I}^{t+1|t}(\mathbf{p}, \bar{\boldsymbol{S}}^{(i)}), \\ \text{subject to} & \mathbf{p} \in \mathscr{C}_{r}(t) \end{array}$$
(57)

which may be solved by enumeration. The optimal solution of (57) corresponds to the selected relay at t + 1.

Note that, following [22], it may be argued that exactly the same set of scenarios may be used by all relays, in all clusters,

and even at all time slots. This, of course, keeps the sampling requirements at a bare minimum, *networkwide*.

In regard to the *per cluster* computational complexity of the proposed SAA-based scheme, at each time slot t, the conditional statistics of the term  $z_{\tau_r}^{f,g}(\cdot, t+1)$  relative to  $m_{\tau_r}^{1:t}$ need to be evaluated, for all  $r = 1, \ldots, N_c$ . As explained above, for a single relay of a specific cluster, this requires an order of  $\mathcal{O}(t^2)$  operations at time slot t, but since this needs to be done for all relays of the cluster, we end up with an order of  $\mathcal{O}(\delta t^2)$  operations. Additionally, for each of the  $|S_r^f \cup S_r^g|$ segments (not containing a cluster) associated with cluster r,  $\mathcal{O}(1)$  complexity is required due to the recursive form of the Kalman filter. Therefore, the total computational complexity of the SAA-scheme for cluster r, in time slot t, is *at most* of the order of  $O(\delta t^2 + |S_r^f \cup S_r^g|)$ ; in fact, in most cases, this complexity is often much smaller, since each of the segments contained in  $S_r^f \cup S_r^g$  may also be associated with clusters other than r, as well.

#### C. 2-STAGE JOINT BEAMFORMING/RELAY SELECTION

Our 2-stage joint beamforming and relay selection scheme is described in Algorithm 1. At time t, beamforming towards the destination is performed, which corresponds to the 2nd stage problem of time slot t. In this stage, the RSS and phases of the channel aggregates at every cluster representative need to be centrally collected, in order to compute the optimal beamforming weights. Within the same time slot t, and in parallel to beamforming, the 1st stage problem of time slot t+1 is solved, i.e., every cluster individually selects the relay to be used for beamforming in the subsequent time slot. In this stage, in addition to the CSI of the cluster representative at  $\mathbf{p}_r$ , the relay selection process also requires the CSI of the unique segments that comprise the propagation paths to that cluster. This information can be easily acquired via low cost devices, e.g., channel sounders, placed on every street segment, and then sent through the backhaul network to the respective cluster.

In practice, it is sufficient to condition on a window of past time slots, as opposed to the entire observed RSS history. Such an approximation is expected to work well even for a relatively small window size, due to the exponentially decaying structure of the temporal correlation component of the channel model. Moreover, depending on the mmWave channel coherence time, it might be sufficient to follow a two-timescale design, where the beamforming weights would be computed in every time slot but relay selection would be executed over a longer time interval.

#### **IV. OPERATIONAL PHASES OF THE TIME SLOT**

In this section, we discuss how the operations of relay selection and beamforming are scheduled within each time slot, comparatively for the proposed and ideal selection schemes.

In every time slot of the ideal scheme, relay selection is always implemented *before* optimal beamforming; this is simply due to the fact that acquisition of the *current* RSS has to inevitably be performed in the same time slot as

### IEEE Access



FIGURE 3. The average SINR observed in every time slot over 10000 independent trial runs, for all policies, when keeping only (a) two (b) four (c) six clusters of Fig. 1. The horizontal lines indicate the average SINR across all time slots for that respective policy.

beamforming. On the other hand, in the proposed scheme, relay selection at the current time slot is implemented *predictively during the previous time slot*, by efficiently exploiting past RSS observations. As a result, the overhead caused by the relay selection process can be effectively bypassed, and optimal beamforming at the current time slot may be implemented *completely in parallel with* the predictive relay selection affecting the next time slot.

Next, let us look at the CSI estimation requirement of each selection scheme into more detail. In the ideal scheme, the incoming and outgoing channels of every relay for all clusters are initially estimated. This requires estimating  $N_{ideal}$  =  $2\delta N_c$  distinct channels. Channel estimation is initiated by a pilot symbol broadcaster from the source, with every relay of all clusters measuring their RSS. A similar procedure is done for estimating the respective channels towards the destination. On the contrary, our proposed scheme requires estimating the CSI of only the cluster representative, as well as the CSI of every associated segment  $\tau \in S_r^f \cup S_r^g$ , for every cluster  $r = 1, \ldots, N_c$ . Therefore,  $N_{proposed} =$  $2N_c + \bigcup_{r=1}^{N_c} S_r^f \cup S_r^g$  channels have to be estimated, where  $|\cdot|$  denotes the cardinality of a set. In other words, when  $\delta > 1 + |\bigcup_{r=1}^{N_c} S_r^f \cup S_r^g|/2N_c$ , our scheme will *always* require less number of channel estimations.

Compared to the ideal, the proposed relay selection scheme is particularly advantageous in dense network topologies, where, to account for high channel variability, each cluster needs to include a large number of relays, and the number of relays per cluster is relatively larger than the number of segments taking part in the communication. Actually, a dense network is required even if the shadowing variance is low, since this implies weaker channel correlation, due to the now dominant multi-path fading. It is then clear that our proposed scheme requires significantly fewer channels to be estimated than the ideal scheme, which in turn leads to reduced channel estimation overhead.

Note that, while our proposed scheme does incur a computational burden associated with the relay selection process, due to the need for execution of Algorithm 1, the parallelization of relay selection and optimal beamforming in each time slot not only compensates for that burden, but also naturally leads to more consistent ergodic performance, as long as the accuracy of predictive relay selection is adequate. Of course, this pressumes that the required computational complexity can be practically implemented within the duration of each time slot.

#### **V. SIMULATIONS**

We examine the performance of the proposed relay selection scheme using synthetically generated CSI data. For our simulations, we assume the topology and cluster placement of Fig. 1, unless otherwise stated. All segments are of length  $d_{full} =$ 100m. For segments containing a relay cluster, we consider  $\delta = 50$  evenly spaced positions upon which the individual relays are located. We assume a carrier frequency of 28 GHz, with path-loss exponent similar to those reported in [5], [36], i.e.  $\alpha_L = \alpha_N = 2.1$ . The assumed channel parameters are  $\eta^2 = 40, \gamma = 15, \beta = 10m, \Delta = 10$  dB,  $\sigma_{\xi}^2 = 20, \sigma^2 = 1$ , and  $\sigma_D^2 = 1$ . The transmission power (which also includes the directional antenna gain) is  $P_S = 80$  dBm and the total relay power budget is  $P_C = 100$  dBm. The relay clusters are connected to a stable power source.

To assess the long term system performance of our SAAbased system, we compare it against two benchmark relay selection policies. The first is the ideal policy, where, as described previously, the best relay from every cluster is selected and used in beamforming, during the current time slot. This policy provides an upper bound on the performance of any admissible policy. The second benchmark is a nonstrategic, randomized policy, where a relay is chosen uniformly at random from its cluster to be used in the subsequent time slot, irrespective of the observed CSI. For the SAA and randomized policies, we also examine the constrained neighborhood case, where the only candidate relays are those in the 4 positions closest to either one of the two cluster segment edges.

All selection policies were executed over  $N_T = 50$  time slots, with  $N_S = 500$  SAA scenarios. This constitutes one trial of the whole communication task. The QoS achieved in each time slot, averaged over  $10^4$  trials, is shown for all policies in Fig. 3b. Note that assessing the performance of the proposed relay selection policies via averaging is correct and technically justified; see [22] for details. The superiority of the SAA policy is evident, as it achieves almost 7.7dB larger SINR than the randomized policy, while also performing only 2.3dB worse than the ideal policy. This significant result implies that near optimal selection is achieved by exploiting one-step ahead SINR prediction, rather than waiting for the actual future channel values, available at the next time slot.

For the ideal and SAA approach, we also show the histograms of the fraction of times a specific relay was chosen in every time slot. Observing the ideal policy of Fig. 5, it seems that there is a tendency to select relays located near either of the two end points of the cluster, which for relay clusters 2 and 3 is done about 60% of the time. This is more evident in the SAA approach where at every time slot about 99% of the total trials choose either of the end point relays.

Under the same channel and simulation parameters, we varied the number of available clusters. We first considered the case where the topology consisted only of clusters 1 and 2 of Fig. 1, and another were in addition to the four clusters of Fig. 1, two more were placed between intersections 6 and 5, and 4 and 7. The average SINR for all policies of the two and six cluster configurations are plotted in Fig. 3a and 3c, respectively. Notice that, as compared to Fig. 3b, the average SINR achieved for the two cluster configuration performs about 5dB worse for all policies, while for the six cluster configuration the performance is about 3dB better. This implies that the more relay clusters present in the city, the better the overall SINR performance achieved at the destination. Of course, the SINR results presented in Figs. 3a,3b and 3c translate naturally into similar trends in terms of the respective achievable ergodic rates.

In Fig. 4a we also examined a topology where two clusters are placed near the source and two near the destination. This case can be thought of as directly assisting the source and destination, fairly. In reality, the probability of this happening in a city is expected to be low as there will be multiple users the clusters would need to service. Still, one can observe that the SINR performance is practically equivalent to that of Fig. 3c. In other words, the cluster placement problem is equally important to the performance of the overall system, and therefore deserves further investigation.

#### **VI. CONCLUSION**

This paper considered a set of spatially distributed clusters of static relays, cooperatively enhancing mmWave communications in an urban environment. In particular, on a per time slot basis, one relay from each cluster is selected to participate in optimal transmit beamforming towards the destination. In order to decide on the best relay configuration to be used for beamforming, we proposed an adaptive, sampling-based, distributed across clusters and of reduced complexity relay selection scheme, which efficiently exploits shadowing-induced spatiotemporal correlations of the mmWave medium, and can be executed completely in



**FIGURE 4.** (a) A different four cluster topology. (b) The average SINR observed in every time slot for all policies, over 10000 independent trial runs for the above cluster topology. The horizontal lines indicate the average SINR across all time slots for that respective policy.

parallel to optimal beamforming-based communication. The effectiveness of our approach is established via numerical simulations, which corroborate its superiority against a RSSagnostic, purely randomized relay selection policy, as well as its efficiency relative to a reference, resource demanding ideal selection scheme. As a byproduct of our results, we also demonstrated the sensitivity of the proposed system with respect to spatial cluster placement, confirming that efficient spatial cluster assignment constitutes an interesting, natural direction for future research.

#### **APPENDIX A**

If we assume the set  $S_r^f = \bigcup_{i=1}^{L_r} S_{ri}^f$  that contains all unique segments of all channels  $f_i$  leading to relay r, as well as the

## IEEE Access



**FIGURE 5.** Ideal policy: fraction of times over 10000 independent trials that a relay was chosen in each time slot, for relay clusters (a) 1 (b) 2 (c) 3 (d) 4 of Fig. 1. For each trial the same relay is initially selected in all clusters, so time slot t = 1 is omitted from the histograms.



**FIGURE 6.** SAA policy: fraction of times over 10000 independent trials that a relay was chosen in each time slot, for relay clusters (a) 1 (b) 2 (c) 3 (d) 4 of Fig. 1. For each trial the same relay is initially selected in all clusters, so time slot t = 1 is omitted from the histograms.

sets  $\mathcal{Z}_r^f = \{z_\tau\}, \tau \in \mathcal{S}_r^f$  and set  $\varphi_r^f = \{\phi_\tau\}$ , we have that:

$$\begin{split} \mathsf{F}(\mathcal{Z}_{r}^{f},\varphi_{r}^{f}) &= |\mathbf{1}^{\mathsf{T}}\mathbf{f}_{r}|^{2} \\ &= |f_{r1}|^{2} + \ldots + |f_{rL_{r}}|^{2} + f_{r1}f_{r2}^{*} + f_{r1}f_{r3}^{*} + \ldots \\ &= |f_{r1}|^{2} + \ldots + |f_{rL_{r}}|^{2} + |f_{r1}|e^{j\Phi_{r1}}|f_{r2}|e^{-j\Phi_{r2}} + \ldots \\ &= \sum_{i=1}^{L_{r}} |f_{ri}|^{2} + |f_{r1}||f_{r2}|(e^{j(\Phi_{r1} - \Phi_{r2})} + e^{j(\Phi_{r2} - \Phi_{r1})}) + \ldots \\ &= \sum_{i=1}^{L_{r}} \sum_{k=1}^{L_{r}} e^{\frac{\chi}{2}F_{ri}}e^{\frac{\chi}{2}F_{rk}}\cos(\Phi_{ri}^{f} - \Phi_{rk}^{f}) \\ &= \sum_{i=1}^{L_{r}} \sum_{k=1}^{L_{r}} \exp\left[\frac{\chi}{2}\left(a_{ri}^{f}(\mathbf{p}_{i}) + z_{\tau_{r}}^{f}(\mathbf{p}_{i}, t) + \sum_{\tau \in \mathcal{S}_{ri}^{f}}z_{\tau}(t)\right)\right] \\ &\quad \times \exp\left[\frac{\chi}{2}\left(a_{rk}^{f}(\mathbf{p}_{k}) + z_{\tau_{r}}^{f}(\mathbf{p}_{k}, t) + \sum_{\tau \in \mathcal{S}_{rk}^{f}}z_{\tau}(t)\right)\right] \\ &\quad \times \cos\left(\phi_{\tau_{r}}^{f}(\mathbf{p}_{i}, t) - \phi_{\tau_{r}}^{f}(\mathbf{p}_{k}, t)\right) \\ &\quad + \sum_{\tau \in \mathcal{S}_{ri}^{f}}\phi_{\tau}(t) - \sum_{\tau \in \mathcal{S}_{rk}^{f}}\phi_{\tau}(t)\right) \end{split}$$

where  $\chi = ln(10)/10$ . Note that in the above, the left hand side refers to the unique segments traversed in all paths, while the equation on the right hand side to the ordered segments in a path. A similar expansion holds for  $G(\mathbb{Z}_r^g, \varphi_r^g)$ .

#### **APPENDIX B**

In the following, superscripts " $Z_r$ " refer to amplitude components, and superscripts " $\varphi_r$ " refer to phase components of the involved aggregate channel. Then, for every pair ( $\mathbf{p}, t + 1$ ), the posterior density of  $S_r(\mathbf{p}, t + 1)$  relative to the RSS

history  $C_r(t)$  is given by (note that we symbolically drop argument dependence in the involved densities)

$$\begin{split} p_{\mathbf{S}_{r}(\mathbf{p},t+1)|\mathcal{C}_{r}(t)} &= \frac{p(\mathbf{S}_{r}(\mathbf{p},t+1),\mathcal{C}_{r}(t))}{p_{\mathcal{C}_{r}(t)}} \\ &= \frac{p_{(\mathbf{S}_{r}(\mathbf{p},t+1),\mathcal{C}_{r}(t))}p_{\mathbf{S}_{r}(\mathbf{p},t+1)}^{\varphi_{r}}}{p_{\mathcal{C}_{r}(t)}^{\mathcal{Z}_{r}}} \\ &= \frac{p_{\mathbf{S}_{r}(\mathbf{p},t+1)|\mathcal{C}_{r}(t)}p_{\mathcal{C}_{r}(t)}^{\mathcal{Z}_{r}}p_{\mathbf{S}_{r}(\mathbf{p},t+1)}^{\varphi_{r}}}{p_{\mathcal{C}_{r}(t)}^{\mathcal{Z}_{r}}} \\ &= p_{\mathbf{S}_{r}(\mathbf{p},t+1)|\mathcal{C}_{r}(t)}p_{\mathbf{S}_{r}(\mathbf{p},t+1)}^{\varphi_{r}}, \end{split}$$

and (33) readily follows by independence of the involved shadowing plus multi-path channel components across distinct segments, also implying conditional independence of those components across segments, relative to  $C_r(t)$ .

#### ACKNOWLEDGMENT

The authors would like to thank the editor and the anonymous reviewers for their insightful comments, which substantially improved the quality and clarity of the paper.

#### REFERENCES

- Federal Communications Commission, "Spectrum frontiers report and order and further notice of proposed rulemaking: FCC16-89," 2016.
- [2] George R MacCartney, Junhong Zhang, Shuai Nie, and Theodore S Rappaport, "Path loss models for 5G millimeter wave propagation channels in urban microcells," in Globecom, 2013, pp. 3948–3953.
- [3] Yong Niu, Yong Li, Depeng Jin, Li Su, and Athanasios V Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges," Wireless Networks, vol. 21, no. 8, pp. 2657–2676, 2015.

- [4] Sundeep Rangan, Theodore S Rappaport, and Elza Erkip, "Millimeterwave cellular wireless networks: Potentials and challenges," Proceedings of the IEEE, vol. 102, no. 3, pp. 366–385, 2014.
- [5] Jae-Joon Park, Jinyi Liang, Juyul Lee, Heon-Kook Kwon, Myung-Don Kim, and Bonghyuk Park, "Millimeter-wave channel model parameters for urban microcellular environment based on 28 and 38 GHz measurements," in 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). IEEE, 2016, pp. 1–5.
- [6] Jae-Hyun Lee, Jeong-Sik Choi, Jung-Yong Lee, and Seong-Cheol Kim, "28 GHz millimeter-wave channel models in urban microcell environment using three-dimensional ray tracing," IEEE Antennas and Wireless Propagation Letters, vol. 17, no. 3, pp. 426–429, 2018.
- [7] Aki Karttunen, Andreas F Molisch, Sooyoung Hur, Jeongho Park, and Charlie Jianzhong Zhang, "Spatially consistent street-by-street path loss model for 28-GHz channels in micro cell urban environments," IEEE Transactions on Wireless Communications, vol. 16, no. 11, pp. 7538– 7550, 2017.
- [8] Sudip Biswas, Satyanarayana Vuppala, Jiang Xue, and Tharmalingam Ratnarajah, "On the performance of relay aided millimeter wave networks," IEEE Journal of Selected Topics in Signal Processing, vol. 10, no. 3, pp. 576–588, 2016.
- [9] Hang Zhao, Rimma Mayzus, Shu Sun, Mathew Samimi, Jocelyn K Schulz, Yaniv Azar, Kevin Wang, George N Wong, Felix Gutierrez, and Theodore S Rappaport, "28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York city," in Communications (ICC), 2013 IEEE International Conference on. IEEE, 2013, pp. 5163–5167.
- [10] Andreas F Molisch, Aki Karttunen, Rui Wang, C Umit Bas, Sooyoung Hur, Jeongho Park, and Jianzhong Zhang, "Millimeter-wave channels in urban environments," in IEEE 10th European Conference on Antennas and Propagation, 2016, pp. 1–5.
- [11] Theodore S Rappaport, George R MacCartney, Shu Sun, Hangsong Yan, and Sijia Deng, "Small-scale, local area, and transitional millimeter wave propagation for 5G communications," IEEE Transactions on Antennas and Propagation, vol. 65, no. 12, pp. 6474–6490, 2017.
- [12] Theodore S Rappaport, Shu Sun, Rimma Mayzus, Hang Zhao, Yaniv Azar, Kevin Wang, George N Wong, Jocelyn K Schulz, Mathew Samimi, and Felix Gutierrez Jr, "Millimeter wave mobile communications for 5G cellular: It will work!," IEEE access, vol. 1, no. 1, pp. 335–349, 2013.
- [13] Nima Palizban, Sebastian Szyszkowicz, and Halim Yanikomeroglu, "Automation of millimeter wave network planning for outdoor coverage in dense urban areas using wall-mounted base stations," IEEE Wireless Communications Letters, vol. 6, no. 2, pp. 206–209, 2017.
- [14] Ioannis Mavromatis, Andrea Tassi, Robert J Piechocki, and Andrew Nix, "Efficient Millimeter-Wave Infrastructure Placement for City-Scale ITS," arXiv preprint arXiv:1903.01372, 2019.
- [15] Shuanshuan Wu, Rachad Atat, Nicholas Mastronarde, and Lingjia Liu, "Improving the coverage and spectral efficiency of millimeter-wave cellular networks using device-to-device relays," IEEE Transactions on Communications, vol. 66, no. 5, pp. 2251–2265, 2018.
- [16] Tianyang Bai and Robert W Heath, "Coverage and rate analysis for millimeter-wave cellular networks," IEEE Transactions on Wireless Communications, vol. 14, no. 2, pp. 1100–1114, 2015.
- [17] Steven W Peters, Ali Y Panah, Kien T Truong, and Robert W Heath, "Relay architectures for 3GPP LTE-advanced," EURASIP Journal on Wireless Communications and Networking, vol. 2009, pp. 1, 2009.
- [18] Xingqin Lin and Jeffrey G Andrews, "Connectivity of millimeter wave networks with multi-hop relaying.," IEEE Wireless Commun. Letters, vol. 4, no. 2, pp. 209–212, 2015.
- [19] Xiaoli Sun, Weiwei Yang, Yueming Cai, Zhongwu Xiang, and Xuanxuan Tang, "Secure Transmissions in Millimeter Wave SWIPT UAV-Based Relay Networks," IEEE Wireless Communications Letters, 2019.
- [20] Furqan Jameel, Shurjeel Wyne, Georges Kaddoum, and Trung Q Duong, "A comprehensive survey on cooperative relaying and jamming strategies for physical layer security," IEEE Communications Surveys & Tutorials, 2018.
- [21] Samy S Soliman and Norman C Beaulieu, "Dual-hop vs multihop af relaying systems," in Global Communications Conference (GLOBECOM), 2013 IEEE. IEEE, 2013, pp. 4299–4305.
- [22] Dionysios S Kalogerias and Athina P Petropulu, "Spatially controlled relay beamforming," IEEE Transactions on Signal Processing, vol. 66, no. 24, pp. 6418–6433, 2018.
- [23] Junting Chen, Urbashi Mitra, and David Gesbert, "Optimal UAV relay placement for single user capacity maximization over terrain with obsta-

cles," International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) 2019, 2019.

- [24] Long Zhang, Hui Zhao, Shuai Hou, Zhen Zhao, Haitao Xu, Xiaobo Wu, Qiwu Wu, and Ronghui Zhang, "A survey on 5G millimeter wave communications for UAV-assisted wireless networks," IEEE Access, 05 2019.
- [25] Guanbo Zheng, Cunqing Hua, Rong Zheng, and Qixin Wang, "Toward robust relay placement in 60 GHz mmWave wireless personal area networks with directional antenna," IEEE Transactions on Mobile Computing, vol. 15, no. 3, pp. 762–773, 2015.
- [26] Ning Wei, Xingqin Lin, and Zhongpei Zhang, "Optimal relay probing in millimeter-wave cellular systems with device-to-device relaying," IEEE Transactions on Vehicular Technology, vol. 65, no. 12, pp. 10218–10222, 2016.
- [27] Tabinda Salam, Jin Xu, Xiaofeng Tao, et al., "Capacity enhancement in 60 GHz based D2D networks by relay selection and scheduling," International Journal of Antennas and Propagation, vol. 2015, 2015.
- [28] Esma Turgut and M Cenk Gursoy, "Energy efficiency in relay-assisted mmwave cellular networks," in 2016 IEEE 84th Vehicular Technology Conference (VTC-Fall). IEEE, 2016, pp. 1–5.
- [29] Minh-Thuyen Thi, Thong Huynh, and Won-Joo Hwang, "Proportional selection of mobile relays in millimeter-wave heterogeneous networks," IEEE Access, vol. 6, pp. 16081–16091, 2018.
- [30] Richard J Weiler, Michael Peter, Wilhelm Keusgen, Kei Sakaguchi, and Fabian Undi, "Environment induced shadowing of urban millimeter-wave access links," IEEE Wireless Communications Letters, vol. 5, no. 4, pp. 440–443, 2016.
- [31] Yuyang Wang, Kiran Venugopal, Robert Wendell Heath, and Andreas F Molisch, "MmWave vehicle-to-infrastructure communication: Analysis of urban microcellular networks," IEEE Transactions on Vehicular Technology, 2018.
- [32] Andreas F. Molisch, Aki Karttunen, Sooyoung Hur, Jeongho Park, and Jianzhong Zhang, "Spatially consistent pathloss modeling for millimeterwave channels in urban environments," in 2016 10th European Conference on Antennas and Propagation (EuCAP). Apr. 2016, IEEE.
- [33] "This Telecom Upstart's 5G Tech Gets a Boost From Facebook," https: //www.wired.com/story/telecom-upstarts-5g-tech-gets-boost-facebook/, Last access, October 2019.
- [34] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński, Lectures on stochastic programming: modeling and theory, SIAM, 2009.
- [35] Luciano Ahumada, Erick Carreño, Albert Angles, and Demian Schkolnik, "Shadowing correlation: Empirical results for mm-Wave wireless links in urban street canyons," IEEE Antennas and Wireless Propagation Letters, vol. 17, no. 4, pp. 543–546, 2018.
- [36] George R MacCartney, Mathew K Samimi, and Theodore S Rappaport, "Omnidirectional path loss models in New York City at 28 GHz and 73 GHz," in 2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC). IEEE, 2014, pp. 227–231.
- [37] Robert W Heath, Introduction to Wireless Digital Communication: A Signal Processing Perspective, Prentice Hall, 2017.
- [38] Jeffrey G Andrews, Tianyang Bai, Mandar N Kulkarni, Ahmed Alkhateeb, Abhishek K Gupta, and Robert W Heath, "Modeling and analyzing millimeter wave cellular systems," IEEE Transactions on Communications, vol. 65, no. 1, pp. 403–430, 2016.
- [39] Mehrzad Malmirchegini and Yasamin Mostofi, "On the spatial predictability of communication channels," IEEE Transactions on Wireless Communications, vol. 11, no. 3, pp. 964–978, 2012.
- [40] Gourab Ghatak, Antonio De Domenico, and Marceau Coupechoux, "Relay placement for reliable ranging in cooperative mm-wave systems," IEEE Wireless Communications Letters, 2019.
- [41] Xingjian Ding, Guodong Sun, Deying Li, Yongcai Wang, and Wenping Chen, "Hop-constrained relay node placement in wireless sensor networks," in International Conference on Wireless Algorithms, Systems, and Applications. Springer, 2018, pp. 78–89.
- [42] Veria Havary-Nassab, Shahram Shahbazpanahi, Ali Grami, and Zhi-Quan Luo, "Distributed beamforming for relay networks based on secondorder statistics of the channel state information.," IEEE Trans. Signal Processing, vol. 56, no. 9, pp. 4306–4316, 2008.
- [43] Gan Zheng, Kai-Kit Wong, Arogyaswami Paulraj, and Björn Ottersten, "Collaborative-relay beamforming with perfect CSI: Optimum and distributed implementation," IEEE Signal Processing Letters, vol. 16, no. 4, pp. 257–260, 2009.

### IEEE Access

- [44] "Deutsche Telekom, Ericsson set out to prove mmWave suitable for 'fiber-like' backhaul," https://www.fiercewireless.com/wireless/deutschetelekom-ericsson-set-out-to-prove-mmwave-suitable-for-fiber-likebackhaul/, Last access, October 2019.
- [45] Raghuraman Mudumbai, D Richard Brown Iii, Upamanyu Madhow, and H Vincent Poor, "Distributed transmit beamforming: challenges and recent progress," IEEE Communications Magazine, vol. 47, no. 2, pp. 102–110, 2009.
- [46] Ali A Nasir, Salman Durrani, Hani Mehrpouyan, Steven D Blostein, and Rodney A Kennedy, "Timing and carrier synchronization in wireless communication systems: a survey and classification of research in the last 5 years," EURASIP Journal on Wireless Communications and Networking, vol. 2016, no. 1, pp. 180, 2016.
- [47] Lin Bai, Tian Li, Quan Yu, Jinho Choi, and Wei Zhang, "Cooperative multiuser beamforming in mmwave distributed antenna systems," IEEE Transactions on Vehicular Technology, vol. 67, no. 12, pp. 12394–12397, 2018.
- [48] Leszek Koschel and Andreas Kortke, "Frequency synchronization and phase offset tracking in a real-time 60-GHz CS-OFDM MIMO system," in 2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications-(PIMRC). IEEE, 2012, pp. 2281–2286.
- [49] AC Ulusoy, S Krone, G Liu, A Trasser, F Guderian, B Almeroth, A Barghouthi, M Hellfeld, S Schumann, C Carta, et al., "A 60 GHz multi-Gb/s system demonstrator utilizing analog synchronization and 1-bit data conversion," in 2013 IEEE 13th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems. IEEE, 2013, pp. 99–101.
- [50] Dionysios S Kalogerias and Athina P Petropulu, "Spatially controlled relay beamforming: 2-stage optimal policies," arXiv preprint arXiv:1705.07463, 2017.

...